

P-22/2

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-231238

(43)Date of publication of application : 05.09.1997

(51)Int.Cl.

G06F 17/30

(21)Application number : 08-058391

(71)Applicant : OMRON CORP

(22)Date of filing : 20.02.1996

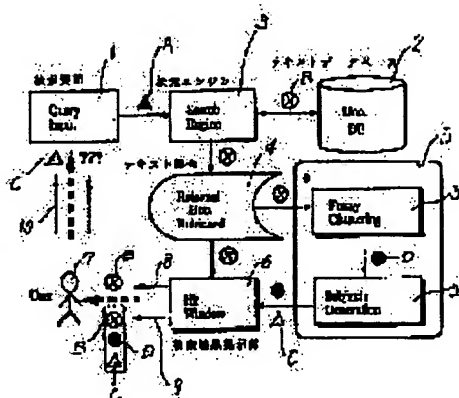
(72)Inventor : GO ATOU
SOGO TAJI
SAWADA AKIRA

(54) DISPLAY METHOD FOR TEXT RETRIEVAL RESULT AND DEVICE THEREFOR

(57)Abstract:

PROBLEM TO BE SOLVED: To improve both retrieval efficiency and accuracy by dividing a text set into plural groups based on the theme analysis result of every text, generates the theme sort information showing the attribute of every group, and displays these information in every group.

SOLUTION: A retrieval engine 3 expands a retrieval expression based on a prescribed algorithm and extracts a relative text set 4 out of a document data base 2. A fuzzy gathering part 5a of a processing part 5 divides the set 4 into plural groups based on the theme analysis result of every text, and a theme sort information generation part 5b generates the theme sort information showing the attribute of every group. A retrieval result display part 6 processes the acquired information (text set B, centroid D and theme sort information C) in a prescribed display mode and shows them to a user 7. As a result, the document retrieval result can be easily confirmed and the retrieval efficiency and accuracy can be improved owing to prevention of the retrieval omission.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

(51)Int.Cl. [*]	識別記号	庁内整理番号	FI	技術表示箇所
G 0 6 F	17/30		G 0 6 F	37 0 A
			15/401	31 0 D
			15/403	38 0 E

特許請求 未請求 請求項の配24 F D (全 19 頁)

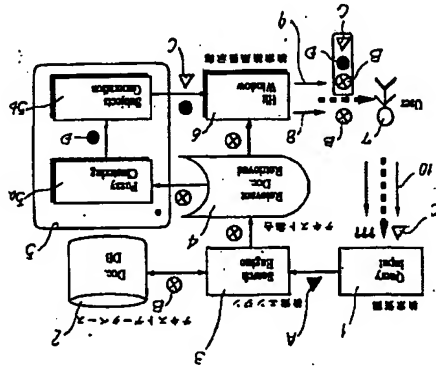
(21) 出願番号	特願平8-58391	(71) 出願人	00002945	才ムロ>株式会社	才
(22) 出願日	平成 8 年 (1996) 2 月 20 日	(72) 発明者	民 重海	京都市京都市右京区花園土堂町10番地	才
		(72) 発明者	十河 太治	京都市京都市右京区花園土堂町10番地	才
			才ムロ>株式会社内	京都市京都市右京区花園土堂町10番地	才
			才ムロ>株式会社内	京都市京都市右京区花園土堂町10番地	才
		(72) 発明者	澤田 晃	京都市京都市右京区花園土堂町10番地	才
			才ムロ>株式会社内	京都市京都市右京区花園土堂町10番地	才
		(74) 代理人	中野士 敏昭 信市	京都市京都市右京区花園土堂町10番地	才

(54) 【発明の名称】 テキスト検索結果表示方法及び装置

(57) 【要約】

【課題】 文書検索結果に対する確認を容易として、検索効率の向上、並びに、検索漏れの防止による検索精度の向上を図ることができ、しかも、提示された主題情報からデータを知問に効率的に絞り定めるかの指示にもなる(Reli. この付加された応答情報を利用して高度な適応検索(Reli. Feedback) を行い得る。

【解決手段】 与えられた検索条件に基づいてデータベースを探索することにより得られたテキスト集合を各テキストの主題分析結果を用いて自動的に複合値のグループに分類し、該分類により得られた各グループのそれぞれについて、当該グループの属性を表設する主題分類情報を生じ、該生成された各グループの主題分類情報をグループ別に区分して表示する。



【特許請求の範囲】

【出典事項1】とえられた検索条件に基づいてデータテーブルを生成することによって得られたテキスト属性を各テキストの主成分分析結果を用いて自動的に複数個のグループに分類する分割ステップと、

【図表例項2】 与えられた検索条件に対してデータベースを逐次探索することによって得られたテキスト集合を各テキストスレッドの主成分分析結果を用いて自動的に該数値のグループに分類する分割ステップによって得られた各グループのそれぞれを生成するステップについて、当該グループの属性を表現する主成分情報（前記各グループのそれらについて、そのグループと前記検索条件との間の重複度を求めるグループ適合度算出ステップで求めた各グループの主成分情報）を、その、前記適合度算出ステップによって求めた適合度の大小を基準として、グループ別に区分して表示する表示ステップを具備する。

ことと特徴とするテキスト検索結果表示方法。

【請求項3】 与えられた検索条件に基づいてデータベースを探索することにより得られたテキスト集合を各テキスト集合の主題分析結果を用いて自動的に数値のグループに分類する分割ステップと、

各グループ内の各テキストの内容の分析結果に基づいて、各テキストの相対グループに対する評価度を算出する評価度算出ステップと、

前記数値のグループの中で、テキスト表示対象となるテキストを選択するための選択ステップと、

前記選択ステップで選択されたグループ内のテキストを、前記選択ステップで選択されたグループ内のテキストと表示する表示ステップとを有する。

【請求項4】 与えられた検索結果表示方法。

【請求項5】 与えられた検索条件に基づいてデータベースを探索することにより得られたテキスト集合を各テキスト集合の主題分析結果を用いて自動的に数値のグループに分類する分割ステップと、

各グループ内の各テキストの内容の分析結果に基づいて、各テキストの相対検索条件に対する評価度を算出する評価度算出ステップと、

前記数値のグループの中で、テキスト表示対象となるテキストを選択するための選択ステップと、

前記選択ステップで選択されたグループ内のテキストを、前記選択ステップで選択されたグループ内のテキストと表示する表示ステップとを有する。

前記算出された適合度の順に内容表示する表示スデッ
とを具備する。

ことを特徴とするテキスト検索結果表示方法。
 [請求項5] 与えられた検索条件に適合してデータベ
 スを構築することにより得られたテキスト集合を各テキ
 ストの主題分析結果を用いて自動的に複数型のグルー
 プに分割する分割ステップと、
 前記グループ内の各テキストの内容の分析結果に基い
 て、各テキストの当該グループに対する所属度を算出す
 る所属度算出ステップと、

前記各グループ内の各テキストの内容の分析結果に基づいて、各テキストの前記検索条件に対する適合度を算出する適合度算出ステップと、
前記検索度のグループの中で、テキスト表示対象となるグループを選択するための適宜表示対象グループ選択ステップと、
前記各グループ内のテキストを検索条件への適合度順に表示するか、或いは当該グループへの所属度の順に表示するかを選択するための表示順序基準選択手段と、
前記表示対象グループ選択ステップで選択されたグループ内のテキストを前記表示対象基準選択手段によって選択された表示順序基準の順に内容を表示する表示ステップとを具備する。

ことを特徴とするテキストは検索結果表示方法。
 【請求項6】 前記配分テーブルは、与えられた検索条件に基づいてデータベースを探索することにより得られたテキスト集合を、フェージング、クラスティング法を用いて複数個のグループに分割する。
 【請求項7】 前記生成プロセスによって生成される当該グループの属性を表現する主成分類似情報は、当該グループの属性を反映するキーワードの組により表すものである。

ことを特徴とする請求項1若しくは請求項2のいずれかに記載のテキスト検索結果表示方法。

【請求項8】 前記生成された分類カテゴリーに属する当該グループの属性を表現する属性分類情報は、当該グループの属性を短い文章により表すものである、ことを特徴とする請求項1若しくは請求項2のいずれかに記載のテキスト検索結果表示方法。

【請求項9】 与えられた検索条件に基いてデータベースを探索することにより得られたテキスト集合の特徴行列に対してファジィ・クラスタリングを行い、各文庫の分類カテゴリーへの所属度を生成する所属度生成ステップと、

前記生成された所属度を用いて、各文庫を1若しくは2以上の分類カテゴリーに割り付ける文庫割り付けステップと、

前記生成された分類カテゴリーの中で、テキスト表示対象

となる分類カテゴリを選択するための分類カテゴリーを選択ステップと、

前記分類カテゴリー選択ステップで選択された分類カテゴリー内のテキストをそのグループに対してする適合度の順に内容表示する表示ステップとを具備する、

ことを特徴とするテキスト検索結果表示方法、

【請求項10】 前記文書割り付けステップは、各文書をその所属度の上位k個の分類カテゴリーに割り付ける、

ことを特徴とする請求項9に記載のテキスト検索結果表示方法、

【請求項11】 前記文書割り付けステップは、各文書のある所属度 α 以上の所属度値を有する分類カテゴリーに割り付ける、

ことを特徴とする請求項10に記載のテキスト検索結果表示方法、

【請求項12】 前記文書割り付けステップは、各文書のカテゴリーの確率分布を考慮して分類カテゴリーに割り付ける、

ことを特徴とする請求項11に記載のテキスト検索結果表示方法、

【請求項13】 与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主成分分析結果を用いて自動的に複数個のグループに分割する分割手段と、

前記分割手段によって得られた各グループのそれぞれについて、当該グループの属性を表現する主成分分析情報を生成する生成手段と、

前記生成手段で求めた各グループの主成分分析情報をグループ別に区分して表示する表示手段とを具備する、

ことを特徴とするテキスト検索結果表示装置、

【請求項14】 与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主成分分析結果を用いて自動的に複数個のグループに分割する分割手段と、

前記分割手段によって得られた各グループのそれぞれについて、当該グループの属性を表現する主成分分析情報を生成する生成手段と、

前記各グループのそれぞれについて、そのグループと前記検索条件との間の適合度を求めるグループ適合度算出手段と、

前記生成手段で求めた各グループの主成分分析情報を、前記適合度算出手段によって求めた適合度の大きい順に、グループ別に区分して表示する表示手段とを具備する、

ことを特徴とするテキスト検索結果表示装置、

【請求項15】 与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主成分分析結果を用いて自動的に複数個のグループに分割する分割手段と、

前記グループ内の各テキストの内容の分析結果に基い

て、各テキストの当該グループに対する所属度を算出する所属度算出手段と、

前記複数個のグループの中で、テキスト表示対象となるグループを選択するための選択手段と、

前記選択手段で選択されたグループ内のテキストを前記算出された所属度の順に内容表示する表示手段とを具備する、

ことを特徴とするテキスト検索結果表示装置、

【請求項16】 与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主成分分析結果を用いて自動的に複数個のグループに分割する分割手段と、

前記グループ内の各テキストの内容の分析結果に基いて、各テキストの前記検索条件に対する適合度を算出する適合度算出手段と、

前記複数個のグループの中で、テキスト表示対象となるグループを選択するための選択手段と、

前記選択手段で選択されたグループ内のテキストを前記算出された適合度の順に内容表示する表示手段とを具備する、

ことを特徴とするテキスト検索結果表示装置、

【請求項17】 与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主成分分析結果を用いて自動的に複数個のグループに分割する分割手段と、

前記グループ内の各テキストの内容の分析結果に基いて、各テキストの当該グループに対する所属度を算出する所属度算出手段と、

前記グループ内の各テキストの内容の分析結果に基いて、各テキストの前記検索条件に対する適合度を算出する適合度算出手段と、

前記複数個のグループの中で、テキスト表示対象となるグループを選択するための表示対象グループ選択手段と、

前記各グループ内のテキストを検索条件への適合度順に表示する表示手段は、或いは当該グループへの所属度の順に表示する表示手段と、

前記表示対象グループ選択手段で選択されたグループ内のテキストを前記表示順序基準選択手段によって選択された表示順序基準の順に内容表示する表示手段とを具備する、

ことを特徴とするテキスト検索結果表示装置、

【請求項18】 前記分割手段は、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を、フジィ・クラスタリング法を用いて複数個のグループに分割する、

ことを特徴とする請求項13乃至請求項17のいずれかに記載のテキスト検索結果表示装置、

【請求項19】 前記生成手段によって生成される当該グループの属性を表現する主成分分析情報は、当該グループの

属性を構つたのキーワードの組により表すものである、ことを特徴とする請求項13若しくは請求項14のいずれかに記載のテキスト検索結果表示装置、

【請求項20】 前記生成手段によって生成される当該グループの属性を表現する主成分分析情報は、当該グループの属性を短い文書により表すものである、

ことを特徴とする請求項13若しくは請求項14のいずれかに記載のテキスト検索結果表示装置、

【請求項21】 与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストに対してフジィ・クラスタリングを行い、各文書毎に各分類カテゴリーへの所属度を生成する所属度生成手段と、

前記生成された所属度を用いて、各文書を1若しくは2以上の分類カテゴリーに割り付ける文書割り付け手段と、

前記複数個の分類カテゴリーの中で、テキスト表示対象となる分類カテゴリーを選択するための分類カテゴリー選択手段と、

前記分類カテゴリー選択手段で選択された分類カテゴリー内のテキストをそのグループに対する適合度の順に内容表示する表示手段とを具備する、

ことを特徴とするテキスト検索結果表示装置、

【請求項22】 前記文書割り付け手段は、各文書をその所属度の上位k個の分類カテゴリーに割り付ける、ことを特徴とする請求項21に記載のテキスト検索結果表示装置、

【請求項23】 前記文書割り付け手段は、各文書をある所属度 α 以上の所属度値を有する分類カテゴリーに割り付ける、

ことを特徴とする請求項21に記載のテキスト検索結果表示装置、

【請求項24】 前記文書割り付け手段は、各文書を力カテゴリーの確率分布を考慮して分類カテゴリーに割り付ける、

ことを特徴とする請求項21に記載のテキスト検索結果表示装置、

【発明の詳細な説明】

【発明の属する技術分野】 この発明は、文書データベースの検索に好適なテキスト検索結果表示方法及び装置に係り、特に、与えられた検索条件に基いてデータベースを検索することにより得られたテキスト集合を各テキストの主成分分析結果を用いて自動的に複数個のグループに分割し、該分割結果により得られた各グループの属性を表現する主成分分析情報をグループ別に区分して表示するようにしたテキスト検索結果表示方法及び装置に関する。

【0002】

【従来の技術】 従来のテキスト検索結果表示装置として、例えば、特開平6-76004号公報に記載された

ものが知られている。

【0003】 この装置は、データベース検索結果を格納するデータベース検索結果格納部と、前記データベース検索結果が有する複数の属性値に利用者の制入力を入力し、各検索結果の距離を算出して検索結果間距離算出手段と、検索結果間距離を用いて検索結果を利用者に指定した属性値は予め定められた個数のグループに分割する検索結果グループ分割手段と、所属グループの重心付近に位置する検索結果を算出するグループ代表検索結果算出手段と、各グループの代表検索結果の中から特定の検索結果を利用者に選択させる代表検索結果選択手段と、代表検索結果が属しているグループ内の全検索結果を表示するグループ内検索結果表示手段とから構成されている。

【0004】 すなわち、この従来の装置においては、non-overlapping手法で構造化された(数値)データベース検索結果をユーザーの指定した分類数に分類するもので、分類されたグループの重心に最も近い検索結果を1件ずつハイロットデータとして表示して利用者に希望するグループを選択させ(順位付けなし)、選択されたグループ内の全検索結果をランキングせずに表示するものである。

【0005】

【発明が解決しようとする課題】 しかしながら、このようない従来の装置においては、(数値)データベース検索結果がグループ内の代表文である場合には、その代表文の内容を随機的に選ぶものが表示され、文書全体が表示されるのでグループの内容が把握し難い。つまり、分類された各グループの主題意味を提示するために、異なるグループの重心に最も近い検索結果を1件ずつハイロットデータとして表示するだけでは、内容的に特定する場がある。むしろ、グループ内の共通的な属性項目群を抽出し、利用に提示することが好ましい。加えて、フルテキスト検索システムの場合にあっては、ハイロットデータとして全ての属性データをそのまま提示することは無意味であり、文書内容を容易に理解できるようにハイロットデータの新しい定義が望まれる。

【0006】 すなわち、このような従来の装置においては、グループ内の重心位置の代表検索結果が表示されるため、代表検索結果がグループ内の代表文である場合には、その代表文の内容を随機的に選ぶものが表示され、文書全体が表示されるのでグループの内容が把握し難い。つまり、分類された各グループの主題意味を提示するために、異なるグループの重心に最も近い検索結果を1件ずつハイロットデータとして表示するだけでは、内容的に特定する場がある。むしろ、グループ内の共通的な属性項目群を抽出し、利用に提示することが好ましい。加えて、フルテキスト検索システムの場合にあっては、ハイロットデータとして全ての属性データをそのまま提示することは無意味であり、文書内容を容易に理解できるようにハイロットデータの新しい定義が望まれる。

【0007】 また、従来の装置においては、グループが検索条件に対する適合度の順に並べられないので、検索目的に合致したグループを選択し難い。加えて、従来の装置においては、グループ内の階層がグループへの所属度の順に並んでいないので、グループの代表検索結果を参照するだけでは、グループのイメージが把握し難い場合でも、他の検索結果を参照してイメージを把握することが困難である。つまり、選択されたグループ内の全検索結果をランキングせずに表示する方式では、分類件数が多くなると、検索結果

果への特定のために利用者の負担が大きくなる。このような負担を軽減して検索効率を向上させるためには、検索結果への特定を促進できるようにランディング機能が望まれる。

【0008】更に、文書は複数の主題を持っているのが通例であるため、一つの文書を一つのクラスにしか分類できない従来の手法では、文書分類結果の表示上では検索結果にズレを生じる虞れがある。そのため、文書検索結果集合に対し主題分類を行う際に複数の異なる（主題を表す）クラスに属することを許すようなoverlapping手法が望まれる。

【0009】この発明は、上述の問題点に鑑みながらなされたものであり、その目的とするところは、文書検索結果に対する確認を容易とし、検索効率の向上、並びに、検索漏れの防止による検索精度の向上を図ることである。しかし、提示された主題情報でデータを如何に効率的に絞り込めるかの指針にもなり、この付加された応答情報を利用して高度な適応検索 (Relavance Feedback) を行い得るようにした検索結果表示方法及び装置を提供することにある。

【0010】

【課題を解決するための手段】この出願の請求項1（又は請求項13）に記載の発明は、与えられた検索条件に基づいてデータベースを検索することにより得られたテキスト集合を各データベースの主題分析結果を用いて自動的に複数のグループに分割する分割ステップ（又は手段）と、前記分割ステップ（又は手段）によって得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報（又は手段）で表す生成ステップ（又は手段）と、前記生成ステップ（又は手段）で表された各グループの主題分類情報をグループ別に区分して表示する表示ステップ（又は手段）とを具備することを特徴とするテキスト検索結果表示方法（又は装置）にある。

【0011】ここで、『データベース』とは、ハードディスクや光ディスク等の大容量記憶媒体に記憶されたテキスト集合やインターネット上に存在するホームページ等のテキスト集合がこれに相当する。

【0012】また、『主題分析』とは、テキストの内容を自動的に示す情報を生成することを意味するものであり、文書内のタイトル上のキーワードの集合を生成するものであっても良い。実施の形態においては、文書を文書空間での特徴ベクトルで表現しているベクトル（F）がこれに相当する。

【0013】また、『主題分類情報』とは、テキストのグループについて、そのグループの内容を端的に示す情報を意味する。実施の形態では、キーワード方式とテキスト方式との2方式が示されている。

【0014】そして、この請求項1（又は請求項13）の発明によれば、グループを端的に表現する情報を付加してグループ別に区分表示するので、検索結果を構成す

るグループの全体像を把握し易くなり、次の処理のためのグループ選択が非常に容易となる。

【0015】この出願の請求項2（又は請求項14）の発明は、与えられた検索条件に基づいてデータベースを検索することにより得られたテキスト集合を各データベースの主題分析結果を用いて自動的に複数のグループに分割する分割ステップ（又は手段）と、前記分割ステップ（又は手段）によって得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を生成する生成ステップ（又は手段）と、前記各グループのそれぞれについて、そのグループと前記検索条件との間の適合度を求めるグループ適合度算出ステップ（又は手段）と、前記生成ステップ（又は手段）で求めた各グループの主題分類情報を、前記適合度算出ステップによって求めた適合度の大きい順に、グループ別に区分して表示する表示ステップ（又は手段）とを具備することとを特徴とするテキスト検索結果表示方法（又は装置）にある。

【0016】そして、この請求項2（又は請求項14）の発明によれば、前記請求項1（又は請求項13）に記載の発明の効果に加えて、検索条件への適合度の順に表示するので、検索目的に合致したグループをグループの内容を確認し易く選択することができる。

【0017】この出願の請求項3（又は請求項15）の発明は、与えられた検索条件に基づいてデータベースを検索することにより得られたテキスト集合を各データベースの主題分析結果を用いて自動的に複数のグループに分割する分割ステップ（又は手段）と、前記グループ内の各テキストの内容の分析結果に基づいて、各テキストの当該グループに対する所屬度を算出する所屬度算出ステップ（又は手段）と、前記複数のグループの中で、テキスト表示対象となるグループを選択するための選択ステップ（又は手段）と、前記選択ステップ（又は手段）で選択されたグループ内のテキストを前記算出された所屬度の順に内容表示する表示ステップ（又は手段）とを具備することとを特徴とするテキスト検索結果表示方法（又は装置）にある。

【0018】そして、この請求項3（又は請求項15）の発明によれば、選択されたグループ内のテキストがグループへの所屬度の順に表示されるので、グループの定数が把握し易くなる。

【0019】この出願の請求項4（又は請求項16）の発明は、与えられた検索条件に基づいてデータベースを検索することにより得られたテキスト集合を各データベースの主題分析結果を用いて自動的に複数のグループに分割する分割ステップ（又は手段）と、前記グループ内の各テキストの内容の分析結果に基づいて、各テキストの当該検索条件に対する適合度を算出する適合度算出ステップ（又は手段）と、前記適合度のグループの中で、テキスト表示対象となるグループを選択するための選択ステップ

（又は手段）と、前記選択ステップ（又は手段）で選択されたグループ内のテキストを前記算出された適合度の順に内容表示する表示ステップ（又は手段）とを具備することとを特徴とするテキスト検索結果表示方法（又は装置）にある。

【0020】そして、この請求項4（又は請求項16）の発明によれば、検索条件に通じたグループを選択し、さらにその中のテキストを検索条件の順に表示するので、検索結果をグループ分けしないでテキストを適合度の順に表示する場合よりも、検索条件に対して適切なテキストが早く確実に表示される。

【0021】この出願の請求項5（又は請求項17）の発明は、与えられた検索条件に基づいてデータベースを検索することにより得られたテキスト集合を各データベースの主題分析結果を用いて自動的に複数のグループに分割する分割ステップ（又は手段）と、前記グループ内の各テキストの内容の分析結果に基づいて、各テキストの当該グループに対する所屬度を算出する所屬度算出ステップ（又は手段）と、前記グループ内の各テキストの内容の分析結果に基づいて、各テキストの前記検索条件に対する適合度を算出する適合度算出ステップ（又は手段）と、前記複数のグループの中で、テキスト表示対象となるグループを選択する前記表示対象グループ選択ステップ（又は手段）と、前記各グループ内のテキストを検索条件への適合度の順に表示するかを選択するための表示順序基準の所屬度算出ステップ（又は手段）と、前記表示対象グループを選択ステップで選択されたグループ内のテキストを前記表示順序基準選択手段にて選択された表示順序基準の順に内容表示する表示ステップ（又は手段）とを具備することとを特徴とするテキスト検索結果表示方法（又は装置）にある。

【0022】そして、この請求項5（又は請求項17）の発明によれば、ユーザーの目的に応じてテキストの表示順序を変えることができる。

【0023】この出願の請求項6（又は請求項18）に記載の発明は、請求項1（又は請求項13）乃至請求項5（又は請求項17）のいずれかに記載のテキスト検索結果表示方法（又は装置）において、前記前記分割ステップ（又は手段）は、与えられた検索条件に基づいてデータベースを検索することにより得られたテキスト集合を、フアンジ・クラスタリング法を用いて複数のグループに分割することとを特徴とするものである。

【0024】そして、この請求項6（又は請求項18）に記載の発明によれば、ある検索式により探し出された文書集合に対して自動的にoverlapping方式で主題内容によるフアンジ分割（主題分類）が行われるため、検索漏れの防止による検索精度の向上が期待できる。

【0025】この出願の請求項7（又は請求項19）に記載の発明は、請求項1（又は請求項13）又は請求項

2（又は請求項14）に記載のテキスト検索結果表示方法（又は装置）において、前記生成ステップ（又は手段）にて生成される当該グループの属性を表現する主題分類情報は、当該グループの属性を保持するのキーワードの組により表すものである。ことを特徴とするものである。

【0026】そして、この請求項7（又は請求項19）に記載の発明によれば、当該グループの属性を保持するのキーワードの組を通して直感的に把握することができ

る。

【0027】この出願の請求項8（又は請求項20）に記載の発明は、請求項1（又は請求項13）又は請求項2（又は請求項14）に記載のテキスト検索結果表示方法（又は装置）において、前記生成ステップ（又は手段）にて生成される当該グループの属性を表現する主題分類情報は、当該グループの属性を短い文章により表すものであることを特徴とするものである。

【0028】そして、この請求項8（又は請求項20）に記載の発明によれば、当該グループの属性を短い文章を通して望んでも判りやすく理解させることができる。

【0029】この出願の請求項9（又は請求項21）に記載の発明は、与えられた検索条件に基づいてデータベースを検索することにより得られたテキスト集合の特徴行列に対してフアンジ・クラスタリング法を行い、各グループに各分類カテゴリへの所屬度を生成する所屬度生成ステップ（又は手段）と、前記生成された所屬度を用いて、各文書割り付けステップ（又は手段）と、前記複数の分類カテゴリの中で、テキスト表示対象となる分類カテゴリを選択するための分類カテゴリ選択ステップ（又は手段）と、前記選択された分類カテゴリ内のテキスト（又は手段）で選択された分類カテゴリ内のテキストをそのグループに対する適合度の順に内容表示する表示ステップ（又は手段）とを具備することとを特徴とするテキスト検索結果表示方法（又は装置）にある。

【0030】そして、この請求項9（又は請求項21）に記載の発明によれば、overlapping手法を用いて各文書を1若しくは2以上の分類カテゴリに割り付け、その状態にて選択された分類カテゴリ内のテキストをそのグループに対する適合度の順に内容表示するため、検索効率の向上、並びに、検索漏れの防止による検索精度の向上を図ることができる。

【0031】この出願の請求項10（又は請求項22）に記載の発明は、前記請求項9（又は請求項21）に記載の発明において、前記文書割り付けステップ（又は手段）は、各文書をその所屬度の上位1個の分類カテゴリに割り付ける。ことを特徴とするものである。

【0032】そして、この請求項10（又は請求項22）に記載の発明によれば、請求項9（又は請求項21）に記載の発明の効果に加え、各分類カテゴリに一つ

いて常に所属度の高い順に一定数順の文 を表示させることができる。

【0033】この出版の請求項11（又は請求項23）に記載の発明は、前記請求項9（又は請求項21）に記載の発明において、前記文書割り付けステップは、各文書のある属値 α 以上の所属度値を有する分類カテゴリに割り付ける、ことを特徴とするものである。

【0034】そして、この請求項11（又は請求項23）に記載の発明によれば、各分類カテゴリについて、前記の発明の効果に加え、各分類カテゴリについて常にある属値 α 以上の所属度値を有する文書を表示させることができる。

【0035】この出版の請求項12（又は請求項24）に記載の発明は、前記請求項9（又は請求項21）に記載の発明において、前記文書割り付けステップは、各文書のカテゴリの確率分布を考慮して分類カテゴリに割り付ける、ことを特徴とするものである。

【0036】そして、この請求項12（又は請求項24）に記載の発明によれば、請求項9（又は請求項21）に記載の発明の効果に加え、各分類カテゴリについてカテゴリーの確率分布を考慮して関連する文書を表示させることができる。

【0037】
【発明の実施の形態】以下に、本発明方法及び装置の好適な実施の形態を添付図面を参照しながら詳細に説明する。

【0038】先ず、本発明方法及び装置が適用されたテキスト検索装置の構成を図1の機能ブロック図により概念的に示す。同図において、1は検索開始時に入力されるべきオリジナル検索質問 (Original Query) や検索絞り込み時に入力されるべきフィードバック検索質問 (Feedback Query) を入力するための検索質問入力部 (Query Input) と記す) であり、具体的には、周知のように、マウスやキーボード等の操作部とそれらの信号を処理する入力用ソフトウェアにより構成される。

【0039】2は検索対象となるテキスト集合に相当するテキスト (文書) データベース (Doc. DB) と記す) であり、具体的には、ハードディスクや光ディスク等の大容量記憶媒体に記憶されたテキスト集合やインターネット上に存在するホームページ等のテキスト集合がこれに相当する。

【0040】3はテキスト検索システムの中核に位置する検索エンジン (Search Engine) と記す) であり、具体的には、周知のように、前述の検索質問入力部1から入力されるオリジナル検索質問 (Original Query) やフィードバック検索質問 (Feedback Query) を検索条件として所定のアルゴリズムに従って検索式を構築し、前述の文書データベース2から関連するテキスト集合を抽出するソフトウェアがこれに相当する。

【0041】4はこのようにして検索エンジン (Search

Engine) 3により抽出された関連するテキスト集合 (Relevant Doc. Retrieved) と記す) であり、後述するように、このテキスト集合4が本発明における加工処理の対象となる。

【0042】5は本発明の要部に関する加工処理部であり、この加工処理部5はテキスト集合4を各グループの主題分析結果を用いて自動的に検索語のグループに分割する分割手段に相当するフuzzy Clustering) と記す) 5 a と、こうして得られた各グループのそれぞれについて、当該グループの属性を表現する主題分類情報を作成する主題分類情報生成部 (Subject Generation) と記す) 5 b とを中心として構成されている。

【0043】フuzzy Clustering) 6 a 及び主題分類情報生成部 (Subject Generation) 5 b の作用を図2に概念的に示す。同図において、符号4で示される実施にて図まれた領域は検索エンジン (Search Engine) 3にて抽出されたテキスト集合 (Retrieved Doc. Retrieved) の全体を表す。

【0044】同様にして、符号4 a、4 b、4 c で示される領域にて図まれた3つの領域はフuzzy Clustering) 6にて分割された3つのグループのそれぞれを表す。

【0045】符号Aで示される黒塗り三角印は、検索開始時に入力されるオリジナル検索質問 (Original Query) を表す。符号Bで示される×入り丸印は、オリジナル検索質問 (Original Query) Aの入力により関連抽出されたテキスト集合4の各構成テキストのそれぞれを表す。

【0046】符号C a、C b、C cで示される3個の白抜き三角印は、グループ4 a、4 b、4 c の属性を表現する主題分類情報 (Group Subject) を表す。尚、これらの主題分類情報C a、C b、C cは検索絞り込みのために用いられ、フィードバック検索質問 (Feedback Query) としても好適なものである。

【0047】符号D a、D b、D cで示される3個の黒塗り丸印は、グループ4 a、4 b、4 c の重心を表す。同様にして、符号Dで示される黒塗り四角印は、テキスト集合4の重心を表す。

【0048】図2から明らかなように、フuzzy Clustering) 6は、検索の結果得られたテキスト集合4に対して、公知のフuzzy Clustering) 6の処理を施すことにより、テキスト集合4を複数個 (この例では3個) のグループ4 a、4 b、4 c に分割する。一方、主題分類情報生成部 (Subject Generation) 5 b は、こうして得られた各グループ4 a、4 b、4 c のそれぞれについて、当該グループの属性を表現する主題分類情報C a、C b、C cを生産する。図から明らかなように、このようにして得られる当該グループの属性を表現する主題分類情報C a、C b、C cは、各グループ4 a、4 b、4 c の重心D a、D b、D cととは異なるもの

であり、まさしくそれぞれのグループの属性を端的に表したものである。尚、これらのフuzzy Clustering) 5 a 及び主題分類情報生成部 (Subject Generation) 5 b の処理内容については、後に、更に詳しく説明する。

【0049】図1に戻って、6は同様に本発明の要部に相当する検索結果提示部 (Hit Window) と記す) であり、この検索結果提示部 (Hit Window) 6では、前述の経過により得られた情報 (テキスト集合B、重心D、主題分類情報C) を所定の表示態様に加工したのち、ユーザ (User) と記す) 7に対して提示する、それらの表示態様についても、後に詳細に説明する。

【0050】尚、図1においては、実施により表された本発明による情報の流れと領域により表された従来装置による情報の流れとが同時に示されている。すなわち、従来装置にあつては、検索結果提示部 (Hit Window) 6では、縦線矢印8に示されるように、検索の結果得られたテキスト集合Bをそのままユーザ7に提示するのみであり、この場合、テキスト集合Bに示されるテキスト数が多量の場合、目的とするテキストを探し出すのにユーザは不便を来す。これに対して、本発明にあつては、検索結果提示部 (Hit Window) 6では、実施矢印9に示されるように、検索の結果得られたテキスト集合Bのみならず、各分類の重心 (Clustercentroids) D並びに主題分類情報 (Group Subject) Cまでもがユーザ7に提示されることとなるため、特に、この主題分類情報 (Group Subject) Cを手軽かりとして、目的とするテキストを容易に探し出すことが可能となる。すなわち、実施矢印10に示されるように、このようにして得られた主題分類情報C (図2のC1、C2、C3に相当する) をそのままフィードバック検索質問 (Feedback Query) Cとして検索質問入力部 (Query Input) 1に与えれば (図2の実施矢印11に検索質問が分岐 "Query Splitting" する様子を示す)、テキスト集合4を的確に絞り込み、目的とするテキストを容易に探し出すことができ、すなわち高度な適応検索 (Relevance Feedback) を行わせることができるのである。

【0051】次に、以上概念的に説明したテキスト検索装置を、さらにその画面表示態様及びそれを実装するためのデータ処理を中心として、図3以下の図面を参照し て詳細に説明する。

【0052】本発明に係るテキスト検索装置におけるデータ処理の全体を図3のゼネラルフローチャートに示す。尚、このゼネラルフローチャートに示される処理は、所定のシステムメニューにおいて、そのメニュー項目のひとつを選択することにより起動される。

【0053】同図において処理が開始されると、検索装置を構成する画像表示部の画面には所定の表示態様に より検索画面が表示される (ステップ301)。このようにして表示される検索画面の一例を図4に示す。同図

に示されるように、表示画面は縦長形状のウィンドウW1により構成されており、その上部略3分の1の部分は検索質問入力領域A1とされており、また下部略3分の2の部分は検索結果出力領域A2とされている。

【0054】検索質問入力領域A1内には検索質問入力用のウィンドウW2が設けられており、このウィンドウW2の上側には、入力ガイド文 (Enter Query in plain English) 12が、またその右側には、前述した検索エンジン (Search Engine) 3に対する起動指示を与えるための起動ボタン (図中OKと記す) 13と、検索質問 (Query) を取り消すための取り消しボタン (図中CA、NCMLと記す) 14と、システムに対して操作支援等を求めるためのヘルプボタン (図中HELPと記す) 15とが設けられている。

【0055】検索結果出力領域A2内には検索結果出力用のウィンドウW3が設けられており、このウィンドウW3の右側にはスクロールバー16が設けられている。更に、この検索結果出力領域A2の下側には、検索結果としてテキスト全文出力を要求するための全文要求ボタン (図中Full Textと記す) 17と、OBJEポタン18と、検索結果の分類化を要求するための分類化要求ボタン (図中Groupingと記す) 19と、検索結果としてテキスト抽出出力を要求するための抽出要求ボタン (図中Summarizeと記す) 20と、画面を検索結果初期化状態に戻すための復帰ボタン (図中Backと記す) 21とが設けられている。

【0056】尚、以上の各々のボタン13、14、15、16、17、18、19、20、21の操作は、カーソルを希望のボタンに移動させた後、マウスのクリック操作等にて行われることは言うまでもない。

【0057】そして、入力ガイド文 (Enter Query in plain English) 12に従って、キーボードから検索質問を自然語 (特に、この例では英語) にて、例えば、"I wantto know Clinton's political condition." の如くに入力すると、この入力された検索質問22はウィンドウW2内に表示されることとなる。

【0058】この状態において、起動ボタン (図中OKと記す) 13が操作されると、図3に戻って、検索装置が実行され、検索エンジン (Search Engine) 3が起動されて、検索質問に関連するテキスト集合4が文書データベース2より抽出され、この抽出されたテキスト集合の各構成テキストは検索質問22との適合度の高い順にソートされ、そのタイトル23のみがウィンドウW3内に表示される (ステップ302)。また、各テキストのタイトル23の先頭部分には、当該テキストの検索質問に対する適合度を三段階 (「高」、「中」、「低」) に区分して表示する適合度マーク24 a、24 b、24 cが表示される。ここで、黒色塗り済みの丸印にて示される適合度マーク24 aは適合度「高」に、灰色塗り済みの丸印にて示される適合度マーク24 bは適合度

【中】に、白抜きの大印にて示される適合度マーク24
oは適合度「低」にそれぞれ対応している。

【0059】以後、図3に示す、システム側において
は文書処理機能の運用を待機する状態となる（ステップ
303）。この状態において、図4の画面に示される分
類化要求ボタン（Grouping）19が操作される
と、本発明の要部である分類化処理が実行される（ステ
ップ306）。

【0060】分類化処理の詳細を図5に示す。図5にお
いて処理が開始されると、所定の案内画面を提示するこ
とにより、分類グループ数gの指定を待機する状態とな
る（ステップ501）。この状態において、分類グルー
プ数gの指定（この例では「5」）が完了すると、本発
明の特徴部分である文 特徴量の抽出処理（ステップ5
02）、フuzzy・クラスタリング処理（Fuzzy Cluste
ringと記す）（ステップ503）、及び主成分分析情報の
生成処理（ステップ504）が順次実行される。

【0061】文書特徴量の抽出処理（ステップ502）
では、次のようにして、文書抽象化と文書特徴ベクトル
を生成要素とするベクトル）によって表され、文書の集
合は語を生成要素とする行列として表される。そのた
め、各文書の特徴となる単語（重要語）を自動的に切り
出し、単語の出現度を決定mとし、各要素が文書単位の単
語の出現頻度と比例するようなベクトル表現F1を用い
ることによって、文書は図1の如くに抽象化される。

【0062】

【数1】
$$F1 = (f11, f12, \dots, f1m)$$

$$F2 = (f21, f22, \dots, f2m)$$

$$\vdots$$

$$Fi = (fi1, fi2, \dots, fim)$$

$$\vdots$$

$$Fj = (fj1, fj2, \dots, fjm)$$

$$\vdots$$

$$Fk = (fk1, fk2, \dots, fkm)$$

$$\vdots$$

$$Fn = (fn1, fn2, \dots, fnm)$$

（例示、或は他の評価値）

数1

文 ベクトル集合の例を表1に示す。この例では、文書
集合の構成文書（F1、F2、F3...）のそれぞれに含
まれる重要語（Clinton, Singapore, China...）の重み
（例えば、頻度）が示されている。

【0063】

【表1】

	Clinton Singapore China ...			
F1	0.8	0.4	0.0	
F2	0.6	0.7	0.0	
F3	0.5	0.0	0.7	
...				

文書ベクトル集合の例

表1

表1に示される文書ベクトル集合を文書空間に展開した
例を図6に示す。この例では、前述の重要語（Clinton,
Singapore, China...）を座標軸とする文書空間に文書集
合の各構成文書（F1、F2、F3...）が展開されてい
る。

【0064】続くフuzzy・クラスタリング処理（ステ
ップ503）では、検索結果としての文書集合の特徴行
列に対し、公知のFCM法を用いてフuzzy・クラスタ
リングを行うことにより、次の2種類の分類情報（V
c、U1）が生成される。

【0065】1） 各分類の代表文書特徴ベクトルVc
【数2】
$$Vc = (vc1, vc2, \dots, vc_m)$$

$$Vg = \text{分類グループ}g \text{の代表文書ベクトル}$$

$$vc_j = \text{分類グループ}j \text{の代表文書}$$

（例示、或は他の重み）

数2

2） 各文書の各分類カテゴリへの所属度U1
【数3】
$$Uj = (u11, u12, \dots, u1g)$$

$$Ui = \text{文書の各分類グループへの所属度ベクトル}$$

$$ui = \text{文書}i \text{の分類グループ}i \text{への所属度}$$

数3

文書分類所属度の例を表2に示す。この例では、各文書
の所属度（U1、U2、U3...）が各分類グループ（G
1、G2、G3...）毎に示されている。

【0066】
【表2】

	G1	G2	...Gg
U1	0.7	0.2	...
U2	0.8	0.1	...
U3	0.3	0.6	...
...			

文書分類所属度の例

表2

続く分類主成分情報の生成処理（ステップ504）では、
次の2種類の方式により、分類主成分情報の生成が行われ
る。

【0067】1） キーワード方式
このキーワード方式は、各分類グループの主題を設つか
のキーワードの組み合わせにより表現する方式であり、
その際、キーワードの抽出には次の2種類の方式が考
えらる。第1の方式は、該当分類の代表文書ベクトル
Vcにおける重みの高い単語の単語を順番にk個抽出し
てそれらの単語をそのグループの主題を表す情報として
用いるものである。第2の方式は、該当分類の文書集合
に対して所属度の高い順にr個の文書ベクトルを抽出
し、そのr個の文書ベクトル集合において出現文書数の
高いものから順にk個の単語を抽出して、そのグルー
プの主題情報を表す情報として用いるものである。

【0068】2） テキスト方式

このテキスト方式では、上記のキーワード方式で主成分
情報を生成するために抽出されたr個の文書の先頭段落の
テキスト（タイトルを含む）に対し、キーワード方式で
得られたキーワード主成分情報を利用して文単位で文字列
集合によりそれらのキーワードを最も多く所有するテキ
ストを抽出し、そのテキスト文をそのグループの主成分
情報として用いるものである。

【0069】このようにして得られた各グループの主題
情報、すなわち分類主成分情報（前述のキーワード又は
タイトル等）は、後述するように、所定の提示順番に
ユーザに提示されることとなる。ここで、検索された
文書1の検索結果に対する適合度をR1、分類グルー
プの検索結果に対する適合度をGrとすると、両者間には数4
の関係が成立する。

【0070】
【数4】

$$Gr = \sum_{i=1}^n Ri / rc$$

Ri: グループiに対して所属度の高い順に抽出された文書数
Ri: グループiに対して抽出された文書数
rc: グループcに対して所属度の高い順に抽出された文書数
rc: グループcに対して抽出された文書数

数4

ここで、数4に示された、グループcに対して所属度の高い順に抽出された文書数rc（c=1, ..., g）は、
分類数の求め方を図7のフローチャートに示す。同図
において、処理が開始されると、rcの初期化（rc=0）
を行ったのち（ステップ701）、文書1の所属度
の付与（U1）に対して最大の所属度が求められ（ステ
ップ702）、その最大値と対応しているグループのメン
バ数rcが加算され（ステップ703）、以上の処
理（ステップ702、703）がiを+1づつ加算しつ
つ（ステップ704）、その加算値がi=n（文書数）
となるまで（ステップ705 YES）繰り返されて、そ
の結果rcの値が最終的に求められることとなる。
【0071】このようにして、分類主成分情報の生成（提
示順番の決定を含む）が完了すると（ステップ50
4）、求められた主成分分類情報を用いた検索結果の動的
表示処理が開始される（ステップ505）。

【0072】検索結果の動的表示処理の詳細を図8のフ
ローチャートに示す。同図において処理が開始される
と、検索結果を構成する画像表示の画面上に設定され
た検索結果出力領域A2は、図9又は図10に示される
ように、上下に2分割され、これにより主成分分類情報
を用いたウィンドウ（Subject Window）W4と検索結果出力
ウィンドウ（Hit Window）W5とが設けられる。そして、
主成分分類情報表示ウィンドウW5（Subject Window）W4
において、所定の表示順番により、各分類主成分情報の
表示が行われる（ステップ801）。前述したように、こ
の各分類主成分情報の提示は、キーワード方式とテキス
ト方式とで実行される。

【0073】キーワード方式による表示画面の一例を図
9に示す。尚、この例では、検索されたテキスト集合が
5個の分類グループに分割されている。同図に示される

ように、主題分類情報表示用ウィンドウ (Subject Window) W4 内には、その左隣部に沿うようにして、分類グループ番号『1』～分類グループ番号『5』に対応する5個のグループボタン25～29が上下一列に配置され、お、それらのグループボタン25～29の右側には、当該分類グループの主題を的確に表すキーワード群30～34が配列されている。この例では、分類グループ番号『1』に対応するグループボタン25の右側には、キーワード群30として、"SINGAPORE:CAFE:PUNIS"が表示されており、分類グループ番号『2』に対応するグループボタン26の右側には、キーワード群31として、"DALAILAMA:MEET:CHINA:TIBET"が表示されており、分類グループ番号『3』に対応するグループボタン27の右側には、キーワード群32として、"WELTLEADER:GOVERNMENT:OFFICIAL"が表示されており、分類グループ番号『4』に対応するグループボタン28の右側には、キーワード群33として、"NIXON:IMMUN:INHAPPORE:DIRECTIVE"が表示されており、分類グループ番号『5』に対応するグループボタン29の右側には、キーワード群34として、"QUESTION:CHARACTER:PEOPLE:POLITICS"が表示されている。

【0074】また、これらの主題分類情報に、先に求められた表示順番に従い、検索質問 (Query) との適合度の高いものから順に配列されている。すなわち、この例では、分類グループ番号『1』にて表される主題が、検索質問との適合度が高く、分類グループ番号『5』にて表される主題が最も検索質問との適合度が低いこととなる。従って、ユーザー7は主題分類情報表示用ウィンドウ (Subject Window) W4 内の表示順番から、自分の望んでいる情報に最も近い分類グループを容易に知ることができ、しかもそれぞれの内容を的確に表すテキスト文35～39の内容に基づいて、各分類グループの主題を大まかに確認することができる。そして、後に詳しく説明するように、分類結果表示処理 (ステップ802) を起動することにより、当初の検索質問に沿うようにして、検索絞り込みを効率よく行うことができる。

【0077】次に、先に説明したフrazier・クラスタリングにより得られた各文書の各分類グループへの所属度U1を用いた、検索結果の最終表示のための処理について詳細に説明する。尚、この例では、分類結果の最終表示のために3種類の処理が用意されており、これらの処理は図9又は図10に示される画面において、グループボタン25～29のいずれか一つを操作することにより起動される (ステップ802)。

【0078】先に説明したように、本発明では検索結果としての文書集合の検索行列に対し、FCM法を用いてフrazier・クラスタリングを行い、それにより各文書の各分類グループへの所属度U1が求められている。今仮に、5個の文書 (001, 002, 003, 004, 005) が存在し、それらの文書のそれぞれについて3個の分類カテゴリ (カテゴリ1、カテゴリ2、カテゴリ3) のそれぞれに対する所属度が表3の通りであると想定する。

【0079】

【表3】

ore Caning, Bulls Response" が表示されており、分類グループ番号『2』に対応するグループボタン26の右側には、テキスト文36として、"Clinton Meets With Dalai Lama" が表示されており、分類グループ番号『3』に対応するグループボタン27の右側には、テキスト文37として、"Indian Leader Meets Clinton" が表示されており、分類グループ番号『4』に対応するグループボタン28の右側には、テキスト文38として、"Nixon Had Living Will" が表示されており、分類グループ番号『5』に対応するグループボタン29の右側には、テキスト文39として、"Clinton News Conference Text" が表示されている。

【0076】また、これらの主題分類情報についても、先に求められた表示順番に従い、検索質問 (Query) との適合度の高いものから順に配列されている。すなわち、この例では、分類グループ番号『1』にて表される主題が、検索質問との適合度が高く、分類グループ番号『5』にて表される主題が最も検索質問との適合度が低いこととなる。従って、ユーザー7は主題分類情報表示用ウィンドウ (Subject Window) W4 内の表示順番から、自分の望んでいる情報に最も近い分類グループを容易に知ることができ、しかもそれぞれの内容を的確に表すテキスト文35～39の内容に基づいて、各分類グループの主題を大まかに確認することができる。そして、後に詳しく説明するように、分類結果表示処理 (ステップ802) を起動することにより、当初の検索質問に沿うようにして、検索絞り込みを効率よく行うことができる。

【0077】次に、先に説明したフrazier・クラスタリングにより得られた各文書の各分類グループへの所属度U1を用いた、検索結果の最終表示のための処理について詳細に説明する。尚、この例では、分類結果の最終表示のために3種類の処理が用意されており、これらの処理は図9又は図10に示される画面において、グループボタン25～29のいずれか一つを操作することにより起動される (ステップ802)。

【0078】先に説明したように、本発明では検索結果としての文書集合の検索行列に対し、FCM法を用いてフrazier・クラスタリングを行い、それにより各文書の各分類グループへの所属度U1が求められている。今仮に、5個の文書 (001, 002, 003, 004, 005) が存在し、それらの文書のそれぞれについて3個の分類カテゴリ (カテゴリ1、カテゴリ2、カテゴリ3) のそれぞれに対する所属度が表3の通りであると想定する。

【表3】

文書番号	カテゴリ1	カテゴリ2	カテゴリ3
001	0.50	0.30	0.20
002	0.60	0.10	0.30
003	0.10	0.80	0.10
004	0.25	0.34	0.41
005	0.35	0.10	0.55

振り付けの説明のための数値例

表3

プログラムの一例を図11に示す。図11において処理が開始されると、k値の設定処理 (ステップ1101) 及び1, c, Ncの初期化処理 (ステップ1102) を実行した後、文書iの所属度データiに対するソート処理 (ステップ1103)、最大所属度データ値から順にk (4)、及び該当するk個のグループに文書iを登録する (ステップ1104) と同時にメンバー数を加算する処理 (ステップ1105) が、文書番号iがnになるまで繰り返され (ステップ1106)、文書番号iがnに達すると各グループ毎の文書振り付け結果を出力して処理が終了 (ステップ1107) する。

【0082】(2) ある閾値α以上の所属度値を有する分類カテゴリに振り付けられる場合
この表示処理にあつては、各文書 (001～005) はある閾値α以上の所属度値を有する分類カテゴリに振り付けられる。ここで、αとしては、例えば1/g (g: 分類数) であることが考えられる。表3に示される例では、g=3, α=0.33となるため、文書 (001) については所属度値が0.33以上であるカテゴリ1に、文書 (002) については同様な理由でカテゴリ1に、文書 (003) については同様な理由でカテゴリ2に、文書 (004) については同様な理由でカテゴリ2とカテゴリ3に、文書 (005) については同様な理由でカテゴリ1とカテゴリ3に振り付けられる。これを分

類カテゴリ (G1, G2, G3) 別に整理すると、

カテゴリG1 = (001, 002, 005) ; N1=3

カテゴリG2 = (003, 004) ; N2=2

カテゴリG3 = (004, 005) ; N3=2

となり、分類グループG1に含まれる文書数N1は3個、分類グループG2に含まれる文書数N2は2個、分類グループG3に含まれる文書数N3は2個とされる。そして、このようにして各カテゴリに属することとされた文書が、後に詳細に説明するように、グループ番号の指定と共に検索結果出力用ウィンドウ (HitWindow) W5内に表示されることとなる。

【0081】以上の表示処理 (1) を実行するためのプ

ロセジューアは、

カテゴリG1 = (001, 002, 005) ; N1=3

カテゴリG2 = (003, 004) ; N2=2

カテゴリG3 = (004, 005) ; N3=2

のグループ番号を抽出する処理 (ステップ1203)、該当する各グループに文書iを登録すると同時にメンバー数を加算する処理 (ステップ1204) が、文書番号iがnになるまで繰り返され (ステップ1205)、文書番号iがnに達すると各グループ毎の文書振り付け結果を出力して処理が終了 (ステップ1206) する。

【0084】(3) カテゴリの確率分布を考慮して分

類カテゴリに振り付けられる場合

プログラムの一例を図12に示す。図12において処理が開始されると、α値の設定処理 (ステップ1201) 及び1, c, Ncの初期化処理 (ステップ1202) を実行した後、文書iの所属度データiに対するu1>α

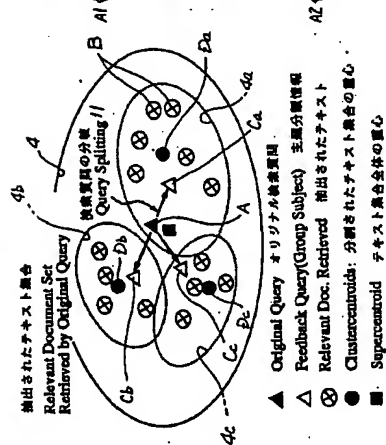
フローチャートである。

【符号の説明】

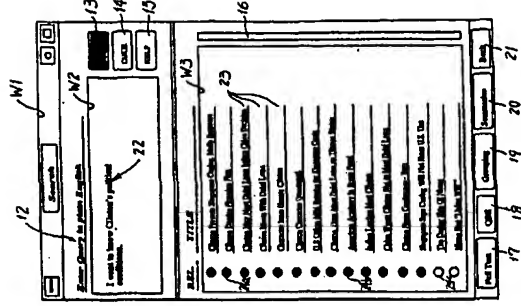
- 1 検索質問入力部
 - 2 文データベース
 - 3 検索エンジン
 - 4 抽出された関連テキスト集合
 - 4a, 4b, 4c 分類グループ
 - 5 加工処理部
 - 5a ファジィ集合化部
 - 5b 主題分類情報生成部
 - 6 検索結果提示部
 - 7 ユーザー
 - 12 入力ガイド文
 - 13 起動ボタン
 - 14 取り換えボタン
 - 15 ヘルプボタン
 - 16 スクロールバー
 - 17 全文要求ボタン
 - 18 QBEボタン
 - 19 分類要求ボタン
 - 20 抄録要求ボタン
 - 21 復帰ボタン
 - 22 検索質問
 - 23 テキスト集合を構成する各テキストのタイトル
- 24a, 24b, 24c 適合度マーク
 - 25~29 グループボタン
 - 30~34 キーワード群
 - 35~39 テキスト文
 - 40~44 テキスト文
 - 45, 46 スクロールバー
 - 49 分類グループ数の表示
 - 47a, 47b, 47c グループ毎の適合度マーク
 - 48a, 48b, 48c 検索質問に対する適合度マーク
 - 49 分類グループ数の表示
 - 50 検索質問適合度順指定ボタン
 - 51 グループ適合度順指定ボタン
 - A オリジナル検索質問
 - B 抽出された各構成テキスト
 - Ca, Cb, Cc 主題分類情報
 - Da, Db, Dc グループの重心
 - A1 検索質問入力領域
 - A2 検索結果出力領域
 - W2 検索質問入力用のウィンドウ
 - W3 検索結果出力用のウィンドウ
 - W4 主題分類情報表示用ウィンドウ
 - W5 検索結果出力用ウィンドウ

トル

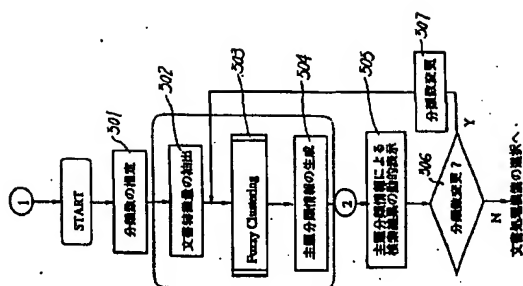
【図2】



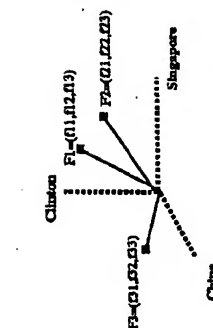
【図4】



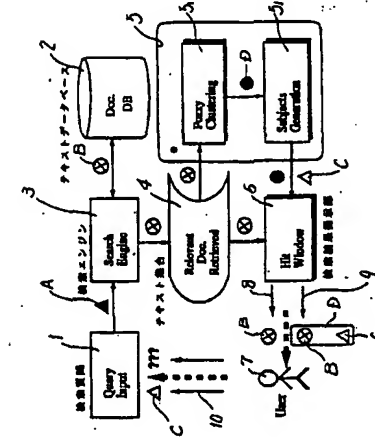
【図5】



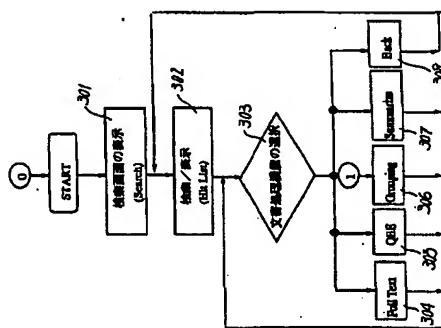
【図6】



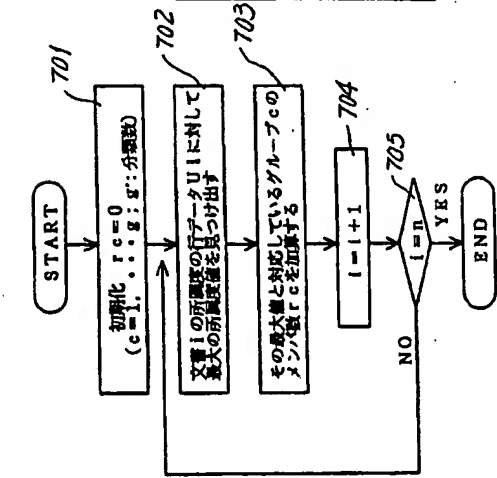
【図11】



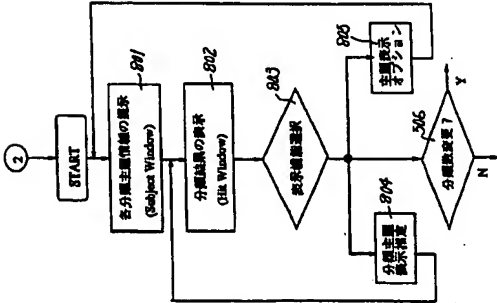
【図3】



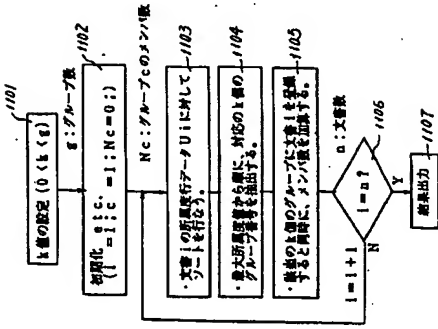
【図 7】



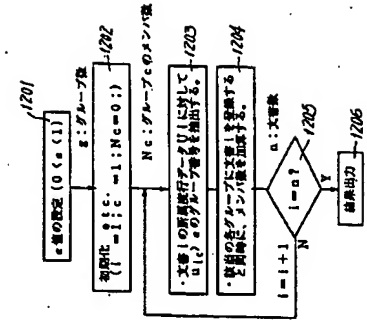
【図 8】



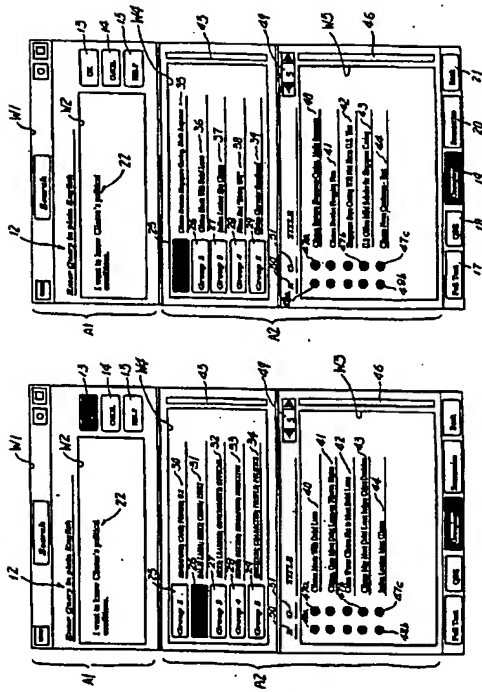
【図 11】



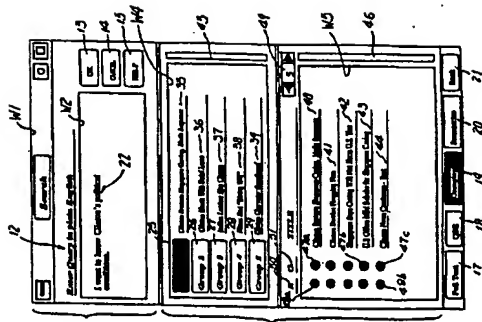
【図 12】



【図 9】



【図 10】



【図13】

